

Implicit U-KAN2.0: Dynamic, Efficient and Interpretable Medical Image Segmentation

Chun-Wun Cheng^{1*}, Yining Zhao^{2*}, Yanqi Cheng¹, Javier A. Montoya-Zegarra^{3,4,5}, Carola-Bibiane Schönlieb¹, and Angelica I Aviles-Rivero⁶(✉)

¹ Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK

² Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign, USA

³ Zurich University of Applied Sciences, Switzerland

⁴ Lucerne Cantonal Hospital, Switzerland

⁵ Lucerne University of Applied Sciences and Arts, Switzerland

⁶ Yau Mathematical Sciences Center, Tsinghua University, China
{aviles-rivero@tsinghua.edu.cn}

Abstract. Image segmentation is a fundamental task in both image analysis and medical applications. State-of-the-art methods predominantly rely on encoder-decoder architectures with a U-shaped design, commonly referred to as U-Net. Recent advancements integrating transformers and MLPs improve performance but still face key limitations, such as poor interpretability, difficulty handling intrinsic noise, and constrained expressiveness due to discrete layer structures, often lacking a solid theoretical foundation. In this work, we introduce Implicit U-KAN 2.0, a novel U-Net variant that adopts a two-phase encoder-decoder structure. In the SONO phase, we use a second-order neural ordinary differential equation (NODEs), called the SONO block, for a more efficient, expressive, and theoretically grounded modeling approach. In the SONO-MultiKAN phase, we integrate the second-order NODEs and MultiKAN layer as the core computational block to enhance interpretability and representation power. Our contributions are threefold. First, U-KAN 2.0 is an implicit deep neural network incorporating MultiKAN and second order NODEs, improving interpretability and performance while reducing computational costs. Second, we provide a theoretical analysis demonstrating that the approximation ability of the MultiKAN block is independent of the input dimension. Third, we conduct extensive experiments on a variety of 2D and a single 3D dataset, demonstrating that our model consistently outperforms existing segmentation networks. Project Website: <https://math-ml-x.github.io/IUKAN2/>

Keywords: Image Segmentation· U-type Networks· Implicit Learning

* Equal contributions.

1 Introduction

Image segmentation is crucial in many applications, particularly in medical imaging and computational sciences, where accurate delineation of anatomical structures is essential. Traditionally, this task has relied on manual annotation by clinicians—a process that is both time-consuming and costly. Recently, deep learning methods, especially convolutional neural networks (CNNs), have significantly improved segmentation accuracy. Among these approaches, U-Net [22] and its variants have become foundational due to their effective encoder-decoder architectures with skip connections. Building on U-Net’s success, numerous extensions have been developed to enhance performance further. For instance, ResUNet [27] incorporated residual units to facilitate the training of deeper networks. Inspired by this work, various block-based architectures [28,13,24] have also been explored. CNN-based models excel at capturing local dependencies but often struggle with global contextual information. To address this, researchers have integrated Transformers into U-Net architectures, with Attention U-Net [19] pioneering the use of attention mechanisms. Later, transformer-based U-Net variants [21,2] were introduced to reduce computational costs while maintaining performance. The Mamba framework [7] emerged as an alternative to traditional Transformers, offering comparable performance with reduced computational complexity from quadratic to linear. Models like U-Mamba [17] and Swin-UMamba [14] further build on the Mamba framework, incorporating aspects of nnU-Net [9] architecture, which requires pre-training. Both CNN-based and transformer-based models discretise continuous functions, while Continuous U-Net [4] offers a continuous block to address this. The continuous formulation of Second Order NODEs [18,3] enables $\mathcal{O}(1)$ memory cost and has been applied in various tasks [25,26,20]. Inspired by Kolmogorov–Arnold Networks (KANs) [16], which use learnable activation functions at edges to optimise feature representation, U-KAN [12] integrates a Tokenised KAN Block with a Convolution Block in U-Net, but relies only on addition. MultiKAN [16] extends KANs by incorporating both addition and multiplication, improving capacity and interpretability.

In this study, we address the limitations of U-KAN by introducing **Implicit U-KAN2.0**, which builds on implicit neural networks for improved efficiency and interpretability. • **Our approach introduces two key novelties:** 1) the *Second Order Neural ODE (SONO) block*, which transforms discrete functions into continuous ones while maintaining constant memory cost, and 2) the *SONO-MultiKAN block*, which integrates SONO with a tokenised MultiKAN layer for enhanced representation power. Compared to U-KAN, our method **redesigns** the architecture by replacing convolutional blocks with SONO, improving discretisation and stability. Furthermore, we introduce a *bottleneck module* to refine the flow of information between the encoder and decoder, optimising feature retention and improving overall model performance. Instead of additive skip connections, we employ feature concatenation to preserve richer representations. Additionally, unlike KAN-based U-Nets, which lack full GPU compatibility, our model is fully optimised for GPU-based training while maintaining constant memory costs and ensuring scalability. • **Our contributions:** I) We introduce a novel implicit

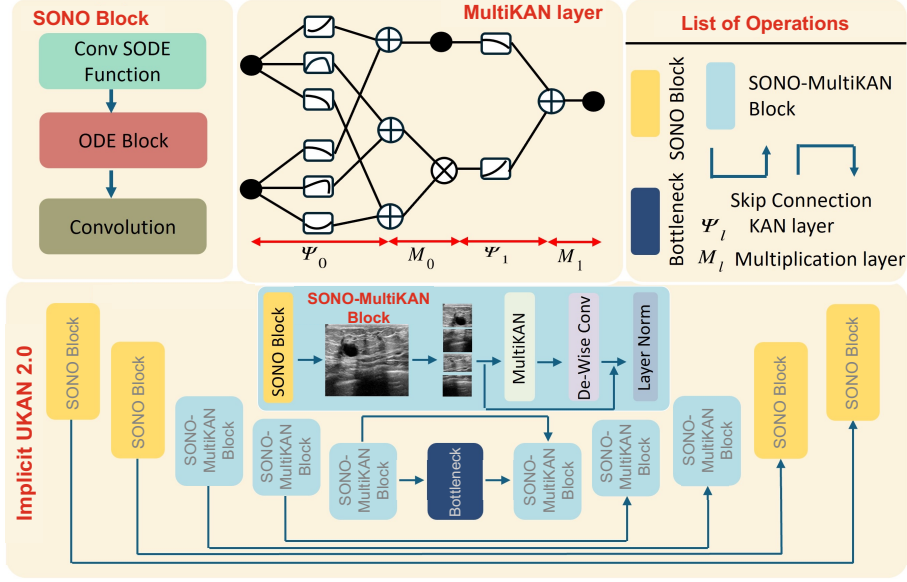


Fig. 1. Overview of Implicit U-KAN2.0. The upper section shows the SONO Block and the MultiKAN layer, while the lower section provides the overall architecture.

deep neural network powered by our SONO block, which enhances the model’s ability to evolve features continuously, improving both accuracy and stability in segmentation tasks. II) We provide theoretical analysis demonstrating that the MultiKAN block’s approximation, with high expressiveness, is independent of input dimensionality. III) Extensive experiments on multiple 2D and 3D datasets confirm that our model outperforms existing segmentation networks.

2 Methodology

This section outlines the key contributions of Implicit U-KAN2.0, highlighting its dynamic feature evolution and enhanced interpretability, which result in higher accuracy and greater efficiency in medical image segmentation tasks. Additionally, we provide a detailed overview of its architecture.

2.1 Implicit U-KAN2.0 is Dynamic

The first key highlight of our Implicit U-KAN2.0 is using Second-Order Neural Ordinary Differential Equations (NODEs) to model the continuous evolution of feature representations, namely SONO block. This dynamic modelling allows for smoother learning trajectories and faster convergence compared to existing methods. The second-order NODEs governing feature evolution is: $\ddot{\mathbf{x}}(t) = f(\mathbf{x}, \dot{\mathbf{x}}, t, \theta_f)$, $\dot{\mathbf{x}}(t_0) = g(x_0, \theta_g)$, $\mathbf{x}(t_0) = x_0$. where f and g are both

neural networks and $\mathbf{x}(t)$ represents the feature vector. The feature was evaluated as a continuous representation and solved by the ode solver in both forward and backward propagation. In addition, the adjoint method was introduced during backpropagation in order to achieve a constant memory cost. By incorporating velocity into the system and extending the solution space to three dimensions, Implicit U-KAN 2.0 achieves faster convergence to optimal feature representations, enhanced stability, and more precise segmentation boundaries. This is particularly important medical imaging where precise boundary delineation is required. To achieve this, we introduce the velocity term $\mathbf{v}(t) = \mathbf{x}'(t)$, reformulating the problem into a system of first-order ordinary differential equations (ODEs): $\dot{\mathbf{x}}(t) = \mathbf{v}(t)$, $\dot{\mathbf{v}}(t) = f(\mathbf{v}(t), t, \boldsymbol{\theta})$. This transformation expands the phase space to $[\mathbf{x}(t), \mathbf{v}(t)]^\top \in \mathbb{R}^{2n}$, which accelerates convergence by enabling trajectory corrections in both position and velocity enabling trajectory corrections in both position and velocity. As a result, convergence is accelerated while maintaining stability. Additionally, the RK4 method is employed to approximate a more stable solution, further improving numerical robustness. This continuous feature evolution ensures smoother transitions in learned representations, which is crucial for medical image segmentation. By capturing gradual changes, it reduces overfitting and makes the network more resilient to noisy, common in medical imaging. This leads to more stable feature extraction and improved segmentation accuracy, even with low-quality images.

2.2 Implicit U-KAN2.0 with higher interpretability

The second key contribution of Implicit U-KAN2.0 involves the integration of Second-order NODEs and MultiKAN layer. Similarly to a SONO block, we first transform the discrete feature vector into a continuous second-order function. Then, the features are further refined using tokenised MultiKAN layer. In KAN, weight matrix was learned by a b-spline function. Inspired by the Kolmogorov-Arnold representation theorem (KART), any continuous function defined in a high-dimensional space can be expressed as a finite combination of continuous functions, each depending on a single variable, along with summation operations. Formally, for any smooth function f , $f(\mathbf{x}) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \varphi_{q,p}(x_p) \right)$, where \mathbf{x} is the input feature vector, $\varphi_{q,p}$ and Φ_q are univariate functions capturing feature interactions. KAN with L layers can be written as a set of composition functions: $\text{KAN}(x) = (\Phi_{L-1} \circ \dots \circ \Phi_1 \circ \Phi_0)(x)$

However, to enhance both interpretability and capacity, MultiKAN was introduced such that incorporates both addition and multiplication operations. MultiKAN extends the standard KANs by interleaving multiplication sub-layers with the conventional addition-based layers. Concretely, the network is defined by two sequences of array sizes: $m^a = [m_0^a, m_1^a, \dots, m_l^a]$ for addition arrays, and $m^n = [m_0^n, m_1^n, \dots, m_l^n]$ for multiplication arrays. A multiplication layer can be divided into two components. The first component transforms the addition nodes without performing any additional actions, while the second component carries out the multiplication. A single MultiKAN layer Ψ_i is constructed by combining

the corresponding KAN layer Φ_i with M_i , and the full MultiKAN is obtained by composing these combined layers.

$$\text{MultiKAN}(x) = (\Psi_L \circ \Psi_{L-1} \circ \cdots \circ \Psi_1)(x). \quad (1)$$

This architecture enriches the interactions among feature arrays and can leverage standard training and regularization methods from the KAN framework because we do not perform additional training in this sub-layer. Similar to KANs, we demonstrated that MultiKAN’s approximation ability is not influenced by dimensionality; rather, it depends on the residual rate.

Theorem 1 (Kolmogorov Arnold theorem for MultiKAN). *Suppose that we define $\text{MultiKAN}(x) = f$ as in (1) and assume $\Psi_{l,i,j}$ is $(k+1)$ -times continuously differentiable. Then, for a given function G -grid B-spline functions, there exist exists a constant C , dependent on f , such that the following holds for all integers m with $0 \leq m \leq k$: $\|\text{MultiKAN}(x) - (\Phi_{L-1}^G \circ \Phi_{L-2}^G \circ \cdots \circ \Phi_1^G \circ \Phi_0^G)x\|_{C^m} \leq C G^{-k-1+m}$.*

Proof. This proof follows a similar approach to that in Theorem 2.1 [16] and we provide a sketch of the proof in here. By the B-spline lemma and the continuity of the functions $\Psi_{l,i,j}$, these functions remain uniformly bounded on any domain. Then we can write $\text{MultiKAN}(x) - (\Phi_{L-1}^G \circ \Phi_{L-2}^G \circ \cdots \circ \Phi_1^G \circ \Phi_0^G)x = Q_{L-1} + Q_{L-2} + \dots + Q_1 + Q_0$, where $Q_l = (\Phi_{L-1}^G \circ \cdots \circ \Phi_{l+1}^G \circ \Phi_l \circ \Phi_{l-1} \circ \cdots \circ \Phi_0)x - (\Phi_{L-1}^G \circ \cdots \circ \Phi_{l+1}^G \circ \Phi_l^G \circ \Phi_{l-1} \circ \cdots \circ \Phi_0)x$ was bounded by CG^{-k-1+m} .

2.3 Implicit UKAN2.0 Architecture

The architecture of Implicit U-KAN2.0 seamlessly integrates its components within a unified encoder-decoder framework. The operation workflow is delineated into two phases: the SONO Phase and the SONO-MultiKAN Phase. Figure 1 shows the overall architecture and the detailed components of each Block.

SONO Phase: The initial input image X_0 is first processed to establish an initial velocity. Subsequently, the extracted features are propagated through a SONO Block, yielding an output defined by: $X_L = \text{Conv}(\text{ODEBlock}(X_{L-1}))$, where $X_\ell \in \mathbb{R}^{H_\ell \times W_\ell \times C_\ell}$, where $X_\ell \in \mathbb{R}^{H_\ell \times W_\ell \times C_\ell}$ denotes the feature map at the l -th layer. Within the SONO Block, we initially employ a function to model the second-order derivative. The resulting output is then fed into an ODE Block, thereby facilitating the approximation of continuous second-order NODEs. Subsequently, we enhance the number of features and perform downsampling via a convolutional layer. In the forward pass, the ODE block is formulated as $\text{ODEBlock}(v_t) = \text{ODESolve}(\mathbf{v}(t_0), f, t_0, t_u, \theta)$ and the loss function is defined by $L(\mathbf{v}(t_u)) = L\left(\mathbf{v}(t_0) + \int_{t_0}^{t_u} f(\mathbf{v}(t), t, \theta) dt\right) = L(\text{ODESolve}(\mathbf{v}(t_0), f, t_0, t_u, \theta))$.

SONO-MultiKAN Phase: In the SONO-MultiKAN phase, the ODE Block is initially applied, followed by the tokenization process to prepare the features for the MultiKAN layer. MultiKAN utilizes learnable activation functions to enhance interpretability. First, the SONO output X_L from the previous two encoders are

transformed into a collection of flattened two-dimensional patches, consistent with the methodology described in [6]. Specifically, X_L is partitioned into M patches, each of dimensions $K \times K$, such that $M = \frac{H_L \times W_L}{K^2}$. Each patch, denoted by $X_L^i \in \mathbb{R}^{K^2 \cdot C_L}$, is subsequently projected into a d -dimensional embedding space through a learnable linear mapping $E \in \mathbb{R}^{(P^2 \cdot C_L) \times d}$. Consequently, the initial token embeddings are represented as $Z_0 = [X_L^1 E; X_L^2 E; \dots; X_L^N E]$. These embeddings are then processed by a MultiKAN module with interleaved multiplicative layers to capture higher-order, non-linear feature interactions. Next, a three-layer MultiKAN module is employed to further extract features, which is then followed by a depth-wise convolutional layer (De-WiseConv). A residual connection is incorporated by adding the original token embeddings, mathematically expressed as: $Z_K = LN(Z_{k-1} + \text{De-WiseConv}(\text{MultiKAN}(\text{ODEBlock}(X_{k-1})))) = LN(Z_{k-1} + \text{De-WiseConv}(\text{MultiKAN}(Z_{k-1})))$. The MultiKAN block provides structural transparency through tokenised basis functions with explicit mathematical roles, unlike saliency maps, which offer approximate explanations for black-box models.

Decoder: Similarly to UNet architecture, the bottleneck was introduced and serves as the critical bridge between the encoder and the decoder. Additionally, skip connections were incorporated, and the decoder comprises two consecutive dynamic blocks and three consecutive MultiKAN blocks for upsampling. We use binary cross-entropy loss for the loss function.

3 Experiments

Datasets and Implementation Detail. We performed experiments using our Implicit U-KAN 2.0 on three distinct 2D medical imaging datasets, namely Kvasir-SEG [10], ISIC Challenge (ISIC) [8] and Breast Ultrasound Images (BU Images) [1]. These datasets encompass a wide range of data types and exhibit substantial variability in terms of image dimensions, segmentation mask fidelity, and overall dataset size. Kvasir-SEG contains 1,000 colonoscopy images (800 for training, 200 for testing) at 256×256 for polyp segmentation. The ISIC Challenge dataset has 1,279 dermoscopy images (990 training, 289 testing) at 512×512 for skin lesion segmentation. The Breast Ultrasound Images dataset includes 647 images (518 training, 129 testing) at 256×256 for breast lesion segmentation. For our 3D experiments, we used the Medical Segmentation Decathlon’s Spleen dataset, which comprises 61 CT volumes (41 training, 20 testing) focusing on the spleen. We implemented the Implicit U-KAN 2.0 model on an NVIDIA A100 40G GPU for testing on the Kvasir-SEG dataset and ISIC Challenge, with batch sizes of 15 and 3 respectively. For the Breast Ultrasound Images, an RTX 4070 Super GPU was used with a batch size of 4. The model was trained for 500 epochs at a learning rate of 0.0001 using early stopping.

3.1 Experiment results and Ablation Studies

• **2D segmentation** In Table 1, we provide a comparison of the performance of different segmentation methods across three datasets. We first compared

Table 1. The comparison of segmentation performance across various methods for three different datasets.

| Dataset | Metric | U-Net | Trans UNet | UNext | USODE | Rolling-UNet | MLLA-UNet | U-KAN | Ours |
|------------|-------------------|--------|------------|--------|--------|---------------|-----------|--------|---------------|
| Kvasir-SEG | Dice \uparrow | 0.7741 | 0.7824 | 0.7583 | 0.7465 | 0.8078 | 0.6962 | 0.7331 | 0.8456 |
| | HD95 \downarrow | 44.75 | 37.97 | 41.12 | 45.20 | 33.74 | 53.62 | 48.40 | 25.26 |
| | Acc. \uparrow | 0.9023 | 0.9064 | 0.8926 | 0.9031 | 0.9278 | 0.8791 | 0.8949 | 0.9134 |
| | IoU \uparrow | 0.4857 | 0.5142 | 0.4657 | 0.4264 | 0.5705 | 0.3574 | 0.4207 | 0.6396 |
| | F1 \uparrow | 0.6074 | 0.6259 | 0.5821 | 0.5506 | 0.6585 | 0.4653 | 0.5289 | 0.7490 |
| ISIC | Dice \uparrow | 0.9091 | 0.9233 | 0.9155 | 0.8998 | 0.9195 | 0.9243 | 0.9190 | 0.9330 |
| | HD95 \downarrow | 15.86 | 10.48 | 14.25 | 24.90 | 9.67 | 10.09 | 12.45 | 7.61 |
| | Acc. \uparrow | 0.9402 | 0.9505 | 0.9448 | 0.9334 | 0.9497 | 0.9514 | 0.9485 | 0.9577 |
| | IoU \uparrow | 0.8044 | 0.8339 | 0.8159 | 0.7854 | 0.8238 | 0.8350 | 0.8244 | 0.8513 |
| | F1 \uparrow | 0.8784 | 0.9009 | 0.8885 | 0.8673 | 0.8934 | 0.9005 | 0.8924 | 0.9128 |
| BU Images | Dice \uparrow | 0.7833 | 0.8117 | 0.7812 | 0.7573 | 0.8187 | 0.7739 | 0.8140 | 0.8397 |
| | HD95 \downarrow | 37.39 | 26.02 | 30.47 | 39.24 | 20.86 | 32.39 | 27.47 | 24.07 |
| | Acc. \uparrow | 0.9498 | 0.9463 | 0.9548 | 0.9433 | 0.9584 | 0.9493 | 0.9587 | 0.9603 |
| | IoU \uparrow | 0.4969 | 0.5673 | 0.4891 | 0.4343 | 0.5762 | 0.4750 | 0.5599 | 0.6132 |
| | F1 \uparrow | 0.5951 | 0.6562 | 0.5882 | 0.5470 | 0.6613 | 0.5768 | 0.6516 | 0.7025 |

with the baseline model U-Net [22], USODE [4] and U-KAN [12]. In addition, we compared with the U-KAN-like structure method UNeXt [23] and Rolling-UNet [15]. We also compared our method with the well-known transformer method TransUNet [21] and state-of-the-art mamba like method MMLA-UNet [11]. We report our performance in terms of Dice, HD95, Accuracy (Acc.), IoU, and F1 Score. Our proposed method consistently outperforms other segmentation models across multiple datasets, with improvements in key metrics such as Dice score, HD95, accuracy, and F1 score. On the Kvasir-SEG dataset, our method achieves a Dice score of 0.8456, representing an improvement of 14.6% to 21.5% over U-KAN (0.7331) and USODE (0.7465), respectively. In terms of HD95, we achieve a 47.7% reduction (from 48.40 to 25.26), demonstrating superior boundary accuracy. We also show improvements in accuracy (up to 5.5% over MMLA-U-Net) and F1 score (41.8% improvement over U-KAN). On BU images, our model achieves a Dice score of 0.8397, improving by 3.2% over U-KAN (0.8140), with the F1 score and IoU also showing improvements ranging from 7.8% to 9.5%.

These results highlight the key strengths of our method: superior segmentation precision and **boundary delineation**. The significant improvements in HD95, particularly a **35.7% reduction** over U-Net, show that our method better captures the fine details of the segmented regions. Additionally, the efficiency of our model, with constant memory costs and GPU-based training, ensures scalability, unlike traditional U-Net-based models that struggle with memory consumption and training time. Overall, our model demonstrates superior performance with **up to 40% improvement** across multiple datasets, making it a **state-of-the-art solution** for medical image segmentation.

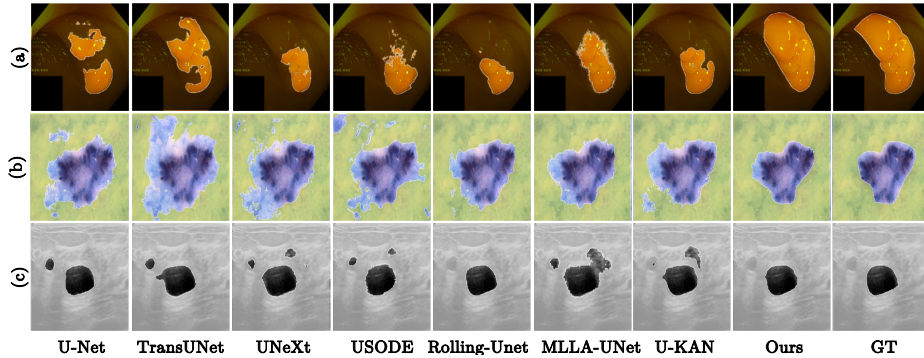


Fig. 2. The visualisation of the 2D segmentation task across the three datasets: (a) Kvasir-SEG [10], (b) ISIC Challenge [8], and (c) Breast Ultrasound Images [1].

Table 2. Comparison of 3D segmentation on Spleen dataset.

| METHOD | Dice |
|----------|--------|
| U-Net 3D | 0.9021 |
| U-KAN 3D | 0.9591 |
| Ours | 0.9687 |

Table 3. Comparison between different noise levels in terms of Dice Socee

| METHOD | ISIC Challenge | | |
|--------|----------------|--------|--------|
| | 0.0 | 0.2 | 0.4 |
| U-KAN | 0.9190 | 0.4064 | 0.4064 |
| Ours | 0.9330 | 0.9225 | 0.9079 |

Figure 2 demonstrates that our method achieves cleaner, more precise segmentations closely matching GT labels across all datasets. While Rolling-U-Net performs well, it shows minor detail inaccuracies. Other methods are less accurate, often missing key structures or creating fragmented regions, underscoring the superior performance of Implicit U-KAN 2.0.

• **3D segmentation.** We implemented both our method and U-KAN in 3D version. From the Table 2, we observe that our proposed Implicit U-KAN 2.0 method significantly outperforms both U-Net 3D [5] and U-KAN 3D. Specifically, our method achieves a Dice score of 0.9687, which is notably higher than the 0.9021 achieved by U-Net 3D and also surpasses U-KAN 3D’s score of 0.9591. This result underscores the superior segmentation performance of our proposed method across different dimensions, highlighting its effectiveness in achieving higher accuracy, as reflected in the Dice score.

• **Ablation Studies.** In the comparison of segmentation performance under different noise levels on the ISIC Challenge dataset, our proposed Implicit U-KAN2.0 outperforms U-KAN at all noise levels. At noise level 0.2, our model achieves a Dice score of 0.9225, while U-KAN drops significantly to 0.4064, demonstrating a 126% improvement. Even at the highest noise level (0.4), our method still delivers a Dice score of 0.9079, compared to U-KAN’s 0.4064, showing a 123% improvement. This highlights the advantage of our approach in handling noisy data, with continuous feature evolution via SONO providing smoother approximations and maintaining robust segmentation despite increased noise. Unlike traditional models, which struggle with noisy data, our model excels in

delivering stable and accurate segmentations, making it particularly suitable for real-world clinical applications where image quality can be compromised.

4 Conclusion

We introduce implicit U-KAN 2.0. It leverages the SONO-Block and SONO-MultiKAN Block, achieving superior accuracy across three benchmark datasets. Our of second-order NODEs blocks enhances both efficiency and noise resistance, while the MultiKAN layer blocks improves interpretability. Additionally, our method outperforms both U-Net and U-KAN in 3D segmentation, highlighting its robustness in handling complex medical image segmentation tasks.

Acknowledgments CWC and JAMZ are supported by the Swiss National Science Foundation under grant number 20HW-1 220785. YC is funded by an AstraZeneca studentship and a Google studentship. CBS acknowledges support from the Philip Leverhulme Prize, the Royal Society Wolfson Fellowship, the EPSRC grants EP/S026045/1, EP/T003553/1, EP/N014588/1, the Wellcome Innovator Award RG98755 and the Alan Turing Institute. AIAR gratefully acknowledges the support of the Yau Mathematical Sciences Center, Tsinghua University. This work is also supported by the Tsinghua University Dushi Program.

Disclosure of Interests The authors declare that they have no competing interests.

References

1. Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. *Data in brief* **28**, 104863 (2020)
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: *European conference on computer vision*. pp. 205–218. Springer (2022)
3. Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. *Advances in neural information processing systems* **31** (2018)
4. Cheng, C.W., Runkel, C., Liu, L., Chan, R.H., Schönlieb, C.B., Aviles-Rivero, A.I.: Continuous u-net: Faster, greater and noiseless. *arXiv preprint arXiv:2302.00626* (2023)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II* 19. pp. 424–432. Springer (2016)
6. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020)
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752* (2023)

8. Gutman, D., Codella, N.C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1605.01397 (2016)
9. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
10. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26. pp. 451–462. Springer (2020)
11. Jiang, Y., Li, Z., Chen, X., Xie, H., Cai, J.: Mlla-unet: Mamba-like linear attention in an efficient u-shape model for medical image segmentation. arXiv preprint arXiv:2410.23738 (2024)
12. Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y.: U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918 (2024)
13. Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.W., Heng, P.A.: H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. IEEE transactions on medical imaging **37**(12), 2663–2674 (2018)
14. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., et al.: Swin-umamba: Mamba-based unet with imagenet-based pretraining. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 615–625. Springer (2024)
15. Liu, Y., Zhu, H., Liu, M., Yu, H., Chen, Z., Gao, J.: Rolling-unet: Revitalizing mlp’s ability to efficiently extract long-distance dependencies for medical image segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 3819–3827 (2024)
16. Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T.Y., Tegmark, M.: Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756 (2024)
17. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
18. Norcliffe, A., Bodnar, C., Day, B., Simidjievski, N., Liò, P.: On second order behaviour in augmented neural odes. Advances in neural information processing systems **33**, 5911–5921 (2020)
19. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
20. Ordoñez, S.C., Cheng, C.W., Huang, J., Zhang, L., Yang, G., Schönlieb, C.B., Aviles-Rivero, A.I.: The missing u for efficient diffusion models. Transactions on Machine Learning Research
21. Petit, O., Thome, N., Rambour, C., Themyr, L., Collins, T., Soler, L.: U-net transformer: Self and cross attention for medical image segmentation. In: Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12. pp. 267–276. Springer (2021)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)

23. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: International conference on medical image computing and computer-assisted intervention. pp. 23–33. Springer (2022)
24. Zhang, L., Dai, J., Lu, H., He, Y., Wang, G.: A bi-directional message passing model for salient object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1741–1750 (2018)
25. Zhang, Y., Cheng, C.W., He, J., He, Z., Schönlieb, C.B., Chen, Y., Aviles-Rivero, A.I.: Cross-modal few-shot learning with second-order neural ordinary differential equations. arXiv preprint arXiv:2412.15813 (2024)
26. Zhang, Y., Cheng, C.W., Yu, K., He, Z., Schönlieb, C.B., Aviles-Rivero, A.I.: Node-adapter: Neural ordinary differential equations for better vision-language reasoning. arXiv preprint arXiv:2407.08672 (2024)
27. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749–753 (2018)
28. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid scene parsing network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890 (2017)